

AD-A183 432

INFERENCE FOR THE BINOMIAL N PARAMETER: A BAYES  
EMPIRICAL BAYES APPROACH REVISION(C) WASHINGTON UNIV  
SEATTLE DEPT OF STATISTICS A E RAFTERY JUL 87 TR-85  
N00014-84-C-0169

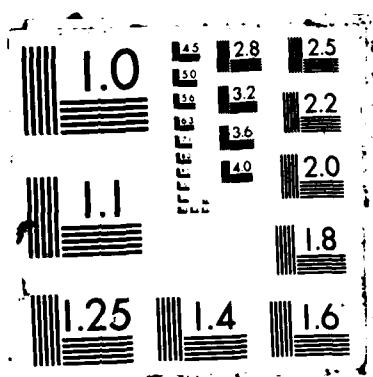
1/1

UNCLASSIFIED

F/G 12/3

NL

END  
9-87  
DTIC



AD-A183 432

12  
DTIC FILE COPY

INFERENCE FOR THE BINOMIAL N PARAMETER:  
A BAYES EMPIRICAL BAYES APPROACH

Adrian E. Raftery

DTIC  
SELECTED  
AUG 14 1987  
S D

TECHNICAL REPORT No. 85

July 1986

REVISED: July 1987

Department of Statistics, GN-22

University of Washington

Seattle, Washington 98195 USA

DISTRIBUTION STATEMENT A  
Approved for public release  
Distribution Unlimited



87724108

A -

## Inference for the Binomial N parameter: A Bayes Empirical Bayes Approach

Adrian E. Raftery

Department of Statistics, GN-22,  
University of Washington,  
Seattle, WA 98195.

### ABSTRACT

The problem of inference about the binomial  $N$  parameter is considered. Applications arise in situations where an unknown population size is to be estimated. Previous work has focused on point estimation, but many applications require interval estimation, prediction, and decision-making.

A Bayes empirical Bayes approach is presented. This provides a simple and flexible way of specifying prior information, and also allows a convenient representation of vague prior knowledge. It yields solutions to the problems of interval estimation, prediction, and decision-making, as well as that of point estimation. The Bayes estimator compares favorably with the best, previously proposed, point estimators in the literature. The Bayesian estimation interval which corresponds to a vague prior distribution also performs satisfactorily when used as a frequentist confidence interval.

A

Distribution For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Government	<input type="checkbox"/>
University	<input type="checkbox"/>
Corporate	<input type="checkbox"/>
By per ltr.	
Classification	
Availability Codes	
Avail. and/or Special	
A-1	

Adrian E. Raftery is Associate Professor of Statistics and Sociology, University of Washington, Seattle, WA 98195. This work was supported by ONR contract N00014-84-C-0169. I am grateful to W.S. Jewell for helpful discussions, and to George Casella, Peter Guttorp, D.V. Lindley, John Petkau, Charles E. Smith, Philip J. Smith, and Philip Turek for helpful comments on an earlier version of this paper.



## 1. INTRODUCTION

Suppose  $x = (x_1, \dots, x_n)$  is a set of success counts from a binomial distribution with unknown parameters  $N$  and  $\theta$ . Most of the literature about statistical analysis of this model has focused on point estimation of  $N$ , which turns out to be a hard problem. This literature is reviewed in Section 3.

However, most of the applications of the model seem to require interval estimation, prediction, or decision-making, which have been little considered in the literature. One common application is animal population size estimation (Carroll and Lombard 1985; Dahiya 1980; Hunter and Griffiths 1978; Moran 1951). Here, presumably, the ultimate purpose of collecting data is to make decisions, such as whether to protect a species which appears to be endangered, or whether to exterminate a pest whose numbers have risen. This may require a full decision-theoretic solution, but often an interval estimate would be sufficient, while a point estimate would not.

Draper and Guttman (1971) used the following application to motivate their work. The  $x_i$  are the numbers of a type of appliance brought in for repair in a service area during week  $i$ , and  $N$  is the number of such appliances in the service area. This seems to be a prediction problem rather than a point estimation one: presumably, the company wants to plan its service facilities, for which it needs to predict future numbers of repairs.

I adopt a Bayes empirical Bayes approach (Deely and Lindley 1981). This provides a simple way of specifying prior information, and also allows a convenient representation of vague prior knowledge using limiting, improper, prior forms. It leads to solutions of the problems of interval estimation, prediction, and decision-making, as well as that of point estimation.

One of the difficulties with Bayesian analysis of this problem has been the absence of a sufficiently flexible and tractable family of prior distributions, mainly due to the fact that  $N$  is an integer. The present, hierarchical, approach gets around this by first assuming that  $N$  has a Poisson distribution. The resulting hyperparameters are then continuous-valued, and one may use existing results about conjugate and vague priors in better understood settings.

A Bayes estimator is shown in Section 3 to compare favorably with the best, previously proposed, point estimators in the literature. The Bayesian estimation interval which corresponds to a vague prior distribution is shown in Section 4 also to perform satisfactorily when used as a frequentist confidence interval.

## 2. A BAYES EMPIRICAL BAYES APPROACH

I assume that  $N$  has a Poisson distribution with mean  $\mu$ . This defines an empirical Bayes model in the sense of Morris (1983). Then each of  $x_1, \dots, x_n$  is a realisation of a Poisson random variable with mean  $\lambda = \mu\theta$ ; the  $x_i$  are not, of course, independent. I carry out a Bayesian analysis of this model.

I specify the prior distribution in terms of  $(\lambda, \theta)$  rather than  $(\mu, \theta)$ . This is because, if the prior is based on past experience, it would seem easier to formulate prior information about  $\lambda$ , the unconditional expectation of the *observations*, than about  $\mu$ , the mean of the *unobserved* quantity  $N$ . For instance, the examples in Section 5 involve estimating the numbers of animals in a National Park based on aerial surveys. Experienced wildlife officials may well have a more precise idea of the number of animals they would see on a particular day, based on the results of previous surveys, than of the number in the entire National Park, which had never been directly observed.

If this is so, the prior information about  $\lambda$  would be more precise than that about  $\mu$  or  $\theta$ , so that it may be more reasonable to assume  $\lambda$  and  $\theta$  independent a priori than  $\mu$  and  $\theta$ . In this case,  $\mu$  and  $\theta$  would be negatively associated a priori. Jewell (1985) has proposed a solution to the different but related problem of population size estimation from capture-recapture sampling, which is based on an assumption similar to prior independence of  $\mu$  and  $\theta$  in the present context.

The posterior distribution of  $N$  is

$$p(N|x) \propto (N!)^{-1} \left\{ \prod_{i=1}^n \left( \frac{N}{x_i} \right) \right\} \int_0^\infty \int_0^\infty \theta^{-N+S} (1-\theta)^{N-S} \lambda^N \exp(-\lambda/\theta) p(\lambda, \theta) d\lambda d\theta$$

(2.1)

$(N \geq x_{\max})$

where  $S = \sum_{i=1}^n x_i$ , and  $x_{\max} = \max\{x_1, \dots, x_n\}$ . If  $\lambda$  and  $\theta$  are independent a priori, and  $\lambda$  has a gamma prior distribution, so that  $p(\lambda, \theta) \propto \lambda^{\kappa_1-1} e^{-\kappa_2 \lambda} p(\theta)$ , then  $\lambda$  can be integrated out analytically, and (2.1) becomes

$$p(N|x) \propto (N!)^{-1} \Gamma(N+\kappa_1) \left\{ \prod_{i=1}^n \left( \frac{N}{x_i} \right) \right\} \int_0^1 \theta^{-N+S} (1-\theta)^{nN-S} (\theta^{-1} + \kappa_2)^{-(N+\kappa_1)} p(\theta) d\theta \quad (N \geq x_{\max})$$

I now consider the case where vague prior knowledge about the model parameters is represented by limiting, improper, prior forms. I use the prior  $p(\lambda, \theta) \propto \lambda^{-1}$ , which is the product of the standard vague prior for  $\lambda$  (Jaynes 1968) with a uniform prior for  $\theta$ . This leads to the same solution as if a similar vague prior were used for  $(\mu, \theta)$ , namely  $p(\mu, \theta) \propto \mu^{-1}$ . It is also equivalent to the prior  $p(N, \theta) \propto N^{-1}$ . The posterior is

$$p(N|x) \propto \{(nN-S)/(nN+1)!N\} \left\{ \prod_{i=1}^n \left( \frac{N}{x_i} \right) \right\} \quad (N \geq x_{\max}) \quad (2.2)$$

The case where  $n = 1$  is of interest as well as of practical importance (Draper and Guttman 1971; Hunter and Griffiths 1978). For example, one may count animals as they migrate past a particular point (Zeh, Ko, Krogman, and Sonntag 1986); inferring the total population from the count is then, in certain situations, an application of the present problem with  $n = 1$ .

When  $n = 1$ , (2.2) becomes

$$p(N|x) = x_1 / \{N(N+1)\} \quad (N \geq x_1)$$

Thus the posterior median is  $2x_1$ . The same solution was obtained by Jeffreys (1961, Section 4.8) to the related problem of estimating the number of bus lines in a town, having seen the number of a single bus. He argued that this was an intuitively reasonable solution, and it seems to be so in this case also.

### 3. POINT ESTIMATION

Most of the literature about the binomial  $N$  problem has focused on point estimation. The problem of estimating  $N$  was first considered by Haldane (1942), who proposed the method of moments estimator, and Fisher (1942), who derived the maximum likelihood estimator. DeRiggi (1983) showed that the relevant likelihood function is unimodal. However, Olkin, Petkau, and Zidek (1981) - hereafter OPZ - showed that both these estimators can be unstable in the sense that a small change in the data can cause a large change in the estimate of  $N$ . Smith and Casella (1986) also report difficulties with maximum likelihood and method of moment estimators of  $N$  for a binomial mixture of normal or gamma random variables, in the context of modeling neurotransmitter release.

OPZ introduced modified estimators and showed that they are stable. On the basis of a simulation study, they recommended the estimator which they called MME:S. Casella (1986) suggested a more refined way of deciding whether or not to use a stabilised estimator. Kappenman (1983) introduced the "sample reuse" estimator; this performed similarly to MME:S in a simulation study, and is not further considered here. Dahiya (1980) used a closely related but different model to estimate the population sizes of different types of organism in a plankton sample by the maximum likelihood method; he did not investigate the stability of his estimators.

Draper and Guttman (1971) adopted a Bayesian approach, and gave a full solution for the case where  $N$  and  $\theta$  are independent a priori, the prior distribution of  $N$  is uniform with a known upper bound, and that of  $\theta$  is beta. Blumenthal and Dahiya (1981) suggested  $N^*$  as an estimator of  $N$ , where  $(N^*, \theta^*)$  is the joint posterior mode of  $(N, \theta)$  with the Draper-Guttman prior. However, they did not say how the parameters of the beta prior for  $\theta$  should be chosen. Carroll and Lombard (1985) - hereafter CL - recommended the  $N$  estimator Mbeta (1,1), the posterior mode of  $N$  with the Draper-Guttman prior after integrating out  $\theta$ , where the prior of  $\theta$  has the form  $p(\theta) \propto \theta(1-\theta)$  ( $0 \leq \theta \leq 1$ ). The Draper-Guttman prior has been criticized by Kahn (1987); see Section 6.

The simpler problem of estimating  $N$  when  $\theta$  is known has been addressed by Feldman and Fox (1968), Hunter and Griffiths (1978), and Sadooghi-Alvandi (1986).

Bayes estimators of  $N$  may be obtained by combining (2.2) with appropriate loss functions; examples are the posterior mode of  $N$ , MOD, and the posterior median of  $N$ , MED. Previous authors, including OPZ, CL, and Casella (1986) have agreed that the relative mean squared error of an estimator  $\hat{N}$ , equal to  $E[(\hat{N}/N-1)^2]$ , is an appropriate loss function for this problem. The Bayes estimator corresponding to this loss function is

$$MRE = \sum_{N=x}^{\infty} N^{-1} p(N|x) / \sum_{N=x}^{\infty} N^{-2} p(N|x)$$

The three Bayes estimators, MOD, MED, and MRE, are reasonably stable, as can be seen from the results for the eight particularly difficult cases listed in Table 2 of OPZ, which are shown in Table 1. MED was closer to the true value of  $N$  than the other estimators considered in four of the eight cases, while MOD was best in a further three cases. However, in the cases in which MOD was best, MED performed poorly; the converse was also true. The other three estimators always fell between MOD and MED.

The results of a Monte Carlo study are shown in Table 2. I used the same design as OPZ and CL. In each replication,  $N$ ,  $\theta$ , and  $n$  were generated from uniform distributions on  $\{1, \dots, 100\}$ ,  $[0,1]$ , and  $\{3, \dots, 22\}$  respectively, using the uniform random number generator of Marsaglia, Ananthanarayanan, and Paul (1973). A binomial success count was then generated using the IMSL routine GGBN. There were 2,000 replications.

Table 2 shows that MRE performed somewhat better than MME:S and Mbeta (1,1) in both stable and unstable cases, with an overall efficiency gain of about 10% over MME:S, and about 6% over Mbeta (1,1). Here, as in OPZ, a sample is defined to be stable if  $\bar{x}/s^2 \geq 1+1/\sqrt{2}$ , and unstable otherwise, where  $\bar{x} = \sum x_i/n$ , and  $s^2 = \sum (x_i - \bar{x})^2/n$ .

Note that, here, MRE is being assessed on the basis of a simulation study designed by the developers of MME:S, where  $N$  is drawn from a distribution very different from, and much

more light-tailed than, the prior for  $N$  on which MRE is based. Presumably, if  $N$  were, instead, simulated from a distribution more similar to the vague prior which leads to (2.2), MRE would perform even better.

*Table 1. N Estimators for Selected and Perturbed Samples.*

Sample	Parameters			Estimators				
	<i>N</i>	$\theta$	<i>n</i>	MME:S	Mbeta (1,1)	MOD	MED	MRE
1	75	.32	5	70	49	42	82	57
				80	52	46	91	62
2	34	.57	4	77	47	42	84	57
				91	52	46	95	62
3	37	.17	20	25	23	21	40	26
				27	25	23	46	29
4	48	.06	15	10	8	7	14	10
				12	10	10	19	12
5	40	.17	12	26	25	23	42	30
				32	29	27	52	35
6	74	.68	12	153	125	114	207	127
				162	131	120	217	129
7	55	.48	20	69	63	59	91	75
				74	67	63	101	81
8	60	.24	15	49	41	38	68	49
				53	45	41	77	53

NOTE: The exact samples are given in Table 2 of OPZ. For each sample number, the first entries are the  $N$  estimates for the original sample, and the second entries are the  $N$  estimates for the perturbed sample obtained by adding one to the largest success count.

*Table 2. Relative Mean Square Errors  
of the N Estimators*

Cases	No.	Estimators		
		MME:S	Mbeta (1,1)	MRE
All cases	2000	.171	.165	.156
Stable cases	1378	.108	.104	.100
Unstable cases	622	.312	.300	.281

#### 4. INTERVAL ESTIMATION

The posterior distribution of  $N$  given by (2.1) or (2.2) yields Bayesian estimation intervals for  $N$ , such as *highest posterior density* (HPD) regions. Such intervals are also exact frequentist confidence intervals in the sense that if the prior distribution also represents a distribution of values of the unknown parameters typical of those that occur in practice, then the average confidence coverage of the Bayesian interval is equal to its posterior probability (Rubin and Schenker 1986). This will not necessarily be the case, however, if the prior distribution used is different from that actually encountered in practice.

In order to evaluate how close the average confidence coverage of HPD regions based on (2.2) is to their posterior probability, a Monte Carlo study, designed in the same way as that reported in Section 3, was carried out. Table 3 shows that the intervals had average confidence coverage close to their posterior probabilities. They are also reasonably stable, as can be seen from Table 4, which shows the intervals for OPZ's eight particularly difficult data sets.

Note that the intervals being evaluated here are based on a prior for  $N$  which is much more diffuse than the, artificially short-tailed, distribution from which  $N$  was simulated. This, together with the asymmetry inherent in the problem, explains the fact that, in Table 4, the true value of  $N$  tends to be closer to the lower than the upper limit of the estimation interval.

To my knowledge, no interval estimators for  $N$ , other than the ones considered here, have been explicitly proposed in the literature. Interval estimators could be constructed based on the Bayesian approach of Draper and Guttman (1971), but they would probably be very sensitive to the, assumed known, prior upper bound for  $N$ , as pointed out by Kahn (1987). Blumenthal and Dahiya (1981, Theorem 5.2 (iii)) did give the asymptotic distribution of the maximum likelihood estimator of  $N$ , and that of their own modified maximum likelihood estimator. While this could, in principle, be used to obtain confidence sets for  $N$ , it did not yield sensible results for many of the real and simulated data sets that I analyzed. Indeed, Blumenthal and Dahiya (1981) did not propose using their result as the basis for a set estimator of  $N$ . A bootstrap interval estimator of  $N$  could be based on any of the proposed point estimators (Efron, 1987), but this possibility has not, so far, been investigated. It would require much more computation than the present approach.

*Table 3. Empirical coverage probabilities of HPD regions*

<i>Posterior probability</i>	<i>Empirical coverage</i>
.80	.82
.90	.91
.95	.95

*Table 4. 80% HPD Regions for Selected and Perturbed Samples*

<i>Sample</i>	<i>N</i>	<i>Limits of 80% HPD region</i>	
		<i>Lower</i>	<i>Upper</i>
1	75	28	211
		30	240
2	34	27	223
		29	258
3	37	13	103
		13	119
4	48	6	35
		7	51
5	40	16	101
		17	131
6	74	72	524
		74	570
7	55	42	181
		43	212
8	60	24	165
		26	191

NOTE: The exact samples are given in Table 2 of OPZ. For each sample number, the first entries are the *N* estimates for the original sample, and the second entries are the *N* estimates for the perturbed sample obtained by adding one to the largest success count.

## 5. EXAMPLES

CL analyzed two examples, involving counts of impala herds and individual waterbuck. The point and interval estimators are shown in Table 3. The stability of the Bayes estimators is again apparent; the stability of MRE for the waterbuck example is noteworthy given the highly unstable nature of this data set.

The posterior distributions obtained from (2.2) are shown in Figures 1 and 2. The posterior distribution for the waterbuck example has a very long tail; this may be related to the extreme instability of this data set.

*Table 5. Point Estimators and 80% HPD regions for the Impala and Waterbuck Examples: Original and Perturbed Samples*

Example	Point Estimators					Limits of 80% HPD region	
	MME:S	Mbeta (1,1)	MOD	MED	MRE	Lower	Upper
Impala	54	42	37	67	49	26	166
	63	46	40	76	54	28	193
Waterbuck	199	140	122	223	131	80	598
	215	146	127	232	132	82	636

NOTE: The data are given in Section 4 of CL. For each example, the first entries are the  $N$  estimates for the original sample, and the second entries are the  $N$  estimates for the perturbed sample obtained by adding one to the largest success count.

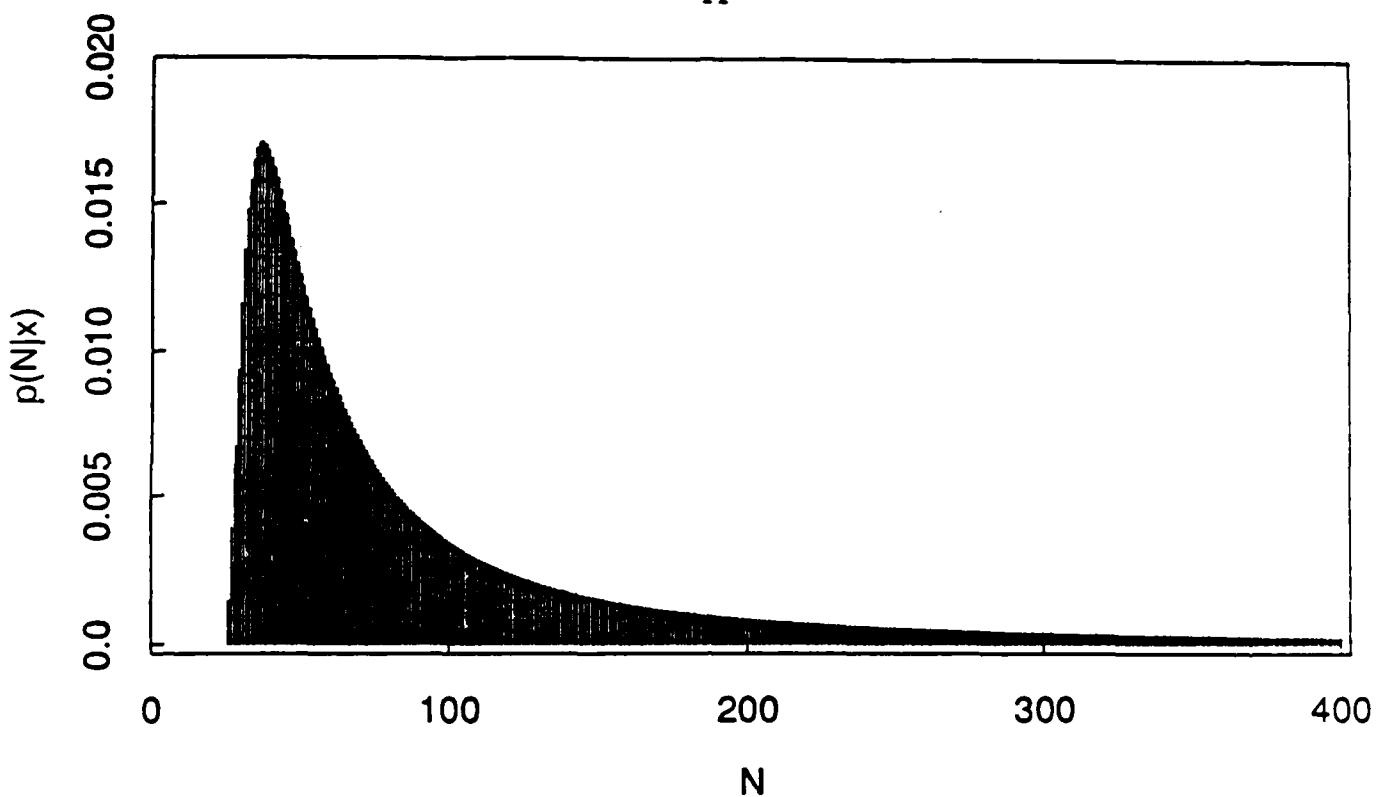


Figure 1. Posterior distribution of  $N$  for the impala example.

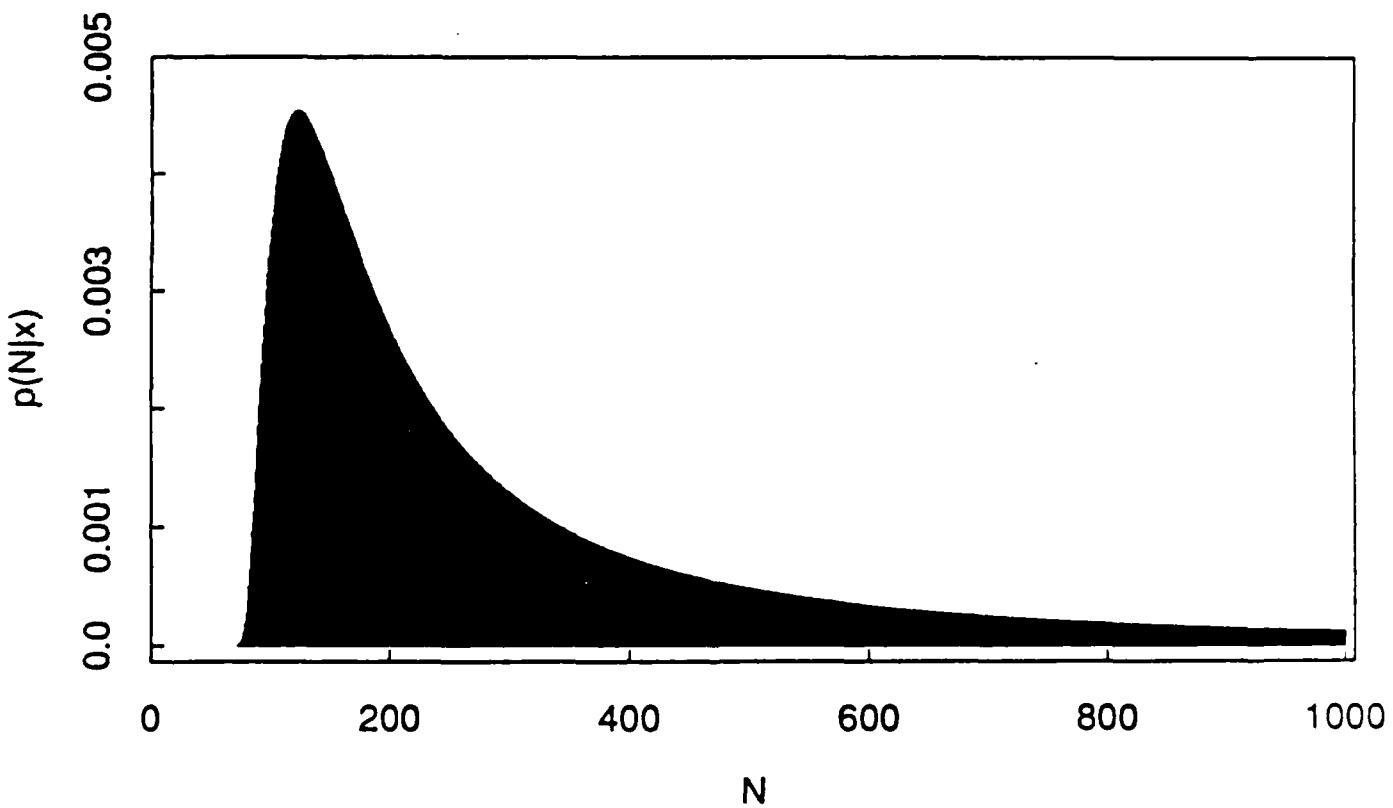


Figure 2. Posterior distribution of  $N$  for the waterbuck example.

## 6. DISCUSSION

I have developed a Bayes empirical Bayes approach to inference about the binomial  $N$  parameter. This provides a simple way of specifying prior information, as well as allowing a convenient representation of vague prior information using limiting, improper, prior forms. It also yields good solutions to the non-Bayesian problems of point and interval estimation. The Bayes point estimator compares favorably with the best, previously proposed, point estimators. The Bayes interval estimator, which currently appears to have no competitors, seems to have about the right average confidence coverage, and to be stable.

The present approach can be used to solve the prediction problem. For example, the predictive distribution of a future observation,  $x_{n+1}$ , is simply

$$p(x_{n+1}|x) \propto \sum_{N=x_{\max}}^{\infty} \int_0^1 p(x_{n+1}, x | N, \theta) p(N, \theta) d\theta$$

When the vague prior which leads to (2.2) is used, this becomes

$$p(x_{n+1}|x) \propto \sum_{N=x_{\max}}^{\infty} \frac{S'! \{(n+1)N - S'\}!}{\{(n+1)N + 1\}! N!} \left\{ \prod_{i=1}^n \binom{N}{x_i} \right\}$$

where  $S' = S + x_{n+1}$  and  $x'_{\max} = \max \{x_{\max}, x_{n+1}\}$ .

No other solution to the prediction problem has, to my knowledge, been explicitly proposed in the literature. A standard, non-Bayesian, approach would be to use the predictive distribution conditional on point estimators of  $N$  and  $\theta$ . As a general method, prediction conditional on the estimated values of the unknown parameters is widespread, and underlies, for example, the time series forecasting methodology of Box and Jenkins (1976). For the present problem, however, it yields predictive distributions which are unsatisfactory because they attribute zero probability to possible outcomes. One consequence of this is that the Kullback-Leibler distance between the true and estimated predictive distributions is often infinite.

My approach also yields a full solution to the decision-making problem, by the usual method of minimizing posterior expected loss. It may often be easier to specify loss (or utility)

in terms of future outcomes than of values of  $N$ , so that a predictive approach to loss specification may be helpful here.

Kahn (1987) has pointed out that in any Bayesian analysis of this problem, the asymptotic tail behaviour of the posterior distribution of  $N$  is determined by the prior. This is not, of course, the same as saying that inferences about  $N$  are determined by the prior. Indeed, in Section 5, we have seen examples where different data lead to very different conclusions about  $N$ , in spite of the priors being the same, and the data sets being small. Kahn (1987) also pointed out that the posterior resulting from the prior used by Draper and Guttman (1971) depends crucially on the, assumed known, prior upper bound for  $N$ , contrary to a comment of Draper and Guttman (1971). The vague prior used here does not appear to suffer from such a problem. Kahn (1987) concluded that the problem should be reparameterized in terms of functions of  $N\theta$  and  $\theta$ , rather than  $N$  and  $\theta$ . This is similar in spirit, if not in technical detail, to the present approach, where I have reparameterized in terms of  $\lambda$  and  $\theta$ , where  $\lambda = E[N]\theta$ . Such a reparameterization alleviates the technical difficulties, and may well, also, make it easier to specify prior information.

## REFERENCES

Blumenthal, S., and Dahiya, R.C. (1981), "Estimating the Binomial Parameter  $n$ ," *Journal of the American Statistical Association*, 76, 903-909.

Box, G.E.P., and Jenkins, G.M. (1976), *Time Series Analysis Forecasting and Control*, (2nd edition), San Francisco: Holden-Day.

Carroll, R.J., and Lombard, F. (1985), "A Note on  $N$  Estimators for the Binomial Distribution," *Journal of the American Statistical Association*, 80, 423-426.

Casella, G. (1986), "Stabilizing Binomial  $n$  Estimators," *Journal of the American Statistical Association*, 81, 172-175.

Dahiya, R.C. (1980), "Estimating the Population Sizes of Different Types of Organisms in a Plankton Sample," *Biometrics*, 36, 437-446.

DeRiggi, D.F. (1983), "Unimodality of Likelihood Functions for the Binomial Distribution," *Journal of the American Statistical Association*, 78, 181-183.

Deely, J.J., and Lindley, D.V. (1981), "Bayes Empirical Bayes," *Journal of the American Statistical Association*, 76, 833-841.

Draper, N., and Guttman, I. (1971), "Bayesian Estimation of the Binomial Parameter," *Technometrics*, 13, 667-673.

Efron, B. (1987), "Better Bootstrap Confidence Intervals (With Discussion)," *Journal of the American Statistical Association*, 82, 171-200.

Feldman, D., and Fox, M. (1968), "Estimation of the Parameter  $n$  in the Binomial Distribution," *Journal of the American Statistical Association*, 63, 150-158.

Fisher, R.A. (1942), "The Negative Binomial Distribution," *Annals of Eugenics*, 11, 182-187.

Haldane, J.B.S. (1942), "The Fitting of Binomial Distributions," *Annals of Eugenics*, 11, 179-181.

Hunter, A.J., and Griffiths, H.J. (1978), "Bayesian Approach to Estimation of Insect Population Size," *Technometrics*, 20, 231-234.

Jaynes, E.T. (1968), "Prior Probabilities," *IEEE Transactions on Systems Science and Cybernetics*, *SSC-4*, 227-241.

Jeffreys, H. (1961), *Theory of Probability*, Oxford: Clarendon Press.

Jewell, W.S. (1985), "Bayesian Estimation of Undetected Errors," *Theory of Reliability*, 94, 405-425.

Kahn, W.D. (1987), "A Cautionary Note for Bayesian Estimation of the Binomial Parameter  $n$ ," *American Statistician*, 41, 38-39.

Kappenman, R.F. (1983), "Parameter Estimation via Sample Reuse," *Journal of Statistical Computation and Simulation*, 16, 213-222.

Marsiglia, G., Ananthanarayanan, K., and Paul, N. (1973), *Super-Duper Random Number Generator*, Montreal: School of Computer Science, McGill University.

Moran, P.A.P. (1951), "A Mathematical Theory of Animal Trapping," *Biometrika*, 38, 307-311.

Morris, C.N. (1983), "Parametric Empirical Bayes Inference: Theory and Applications (with Discussion)," *Journal of the American Statistical Association*, 78, 47-65.

Olkin, I., Petkau, J., and Zidek, J.V. (1981), "A Comparison of  $n$  Estimators for the Binomial Distribution," *Journal of the American Statistical Association*, 76, 637-642.

Rubin, D.B., and Schenker, N. (1986), "Efficiently Simulating the Coverage Properties of Interval Estimates," *Applied Statistics*, 35, 159-167.

Sadooghi-Alvandi, S.M. (1986), "Admissible Estimation of the Binomial Parameter  $n$ ," *Annals of Statistics*, 14, 1634-1641.

Smith, C.E., and Casella, G. (1986), "Some Difficulties With Parameter Estimation for Binomial Neurotransmitter Release," Paper presented at the Biometric Society ENAR Meeting, Atlanta, Georgia, March 1986.

Zeh, J.E., Ko, D., Krogman, B.D., and Sonntag, R. (1986), "A Multinomial Model for Estimating the Size of a Whale Population from Incomplete Census Data," *Biometrics*, 42, 1-14.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 85	2. GOVT ACCESSION NO. AD A103 - 12	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Binomial N estimation: A Bayes empirical Bayes approach	5. TYPE OF REPORT & PERIOD COVERED TR 12/86 - 6/88	
7. AUTHOR(s) Adrian E. Raftery	6. PERFORMING ORG. REPORT NUMBER N00014-84-C-0169	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, GN-22 University of Washington Seattle, WA 98195	10. PROGRAM ELEMENT, PRO. ECT. TASK AREA & WORK UNIT NUMBERS NR-661-003	
11. CONTROLLING OFFICE NAME AND ADDRESS ONR Code N63374 1107 NE 45th Street Seattle, WA 98105	12. REPORT DATE July 1986/REVISED July 1987	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	13. NUMBER OF PAGES 12	
	15. SECURITY CLASS. (of this report) Unclassified	
	16a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report)  APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)  Binomial N estimation; Bayes empirical Bayes		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  A Bayes empirical Bayes approach to the problem of estimating N in the binomial distribution is presented. This provides a simple and flexible way of specifying prior information, and also allows a convenient representation of vague prior knowledge. In addition, it yields a solution to the interval estimation problem. The Bayes estimator corresponding to the relative squared error loss function and a vague prior distribution is shown to be stable, and to compare favorably with the estimators introduced by Olkin et al		

END

9-87

DTIC